

DOCUMENT RESUME

ED 155 271

UD 018 300

AUTHOR Berk, Richard A.; And Others
TITLE Statistical Data in School Court Cases: An Interchange. Reprint 258.
INSTITUTION Wisconsin Univ., Madison. Inst. for Research on Poverty.
PUB DATE 77
NOTE 29p.
EDRS PRICE MF-\$0.83 HC-\$2.06 Plus Postage.
DESCRIPTORS Comparative Analysis; *Court Cases; Educational Finance; Educational Resources; Elementary Education; *Elementary Schools; Equal Education; *Equal Facilities; *Evaluation Methods; School District Spending; Statistical Analysis; *Statistical Data
IDENTIFIERS *Pennsylvania (Philadelphia)

ABSTRACT

This paper details a method for those interested in investigating the equity aspects of intradistrict distributions of educational inputs (financial resources). The method can be viewed as a procedure for post "Hobson v. Hansen" lawyers and for other constituents that departs from some of the speculative quality of the evidence in use. Using detailed individual school data in a systematic way allows the investigator to delineate which unequally distributed inputs positively contribute to stated equity objectives, which inputs are outside the school administration's control, and what the overall equity picture is, starting from the components. An analysis of data from elementary schools in the Philadelphia school district serves as an example of the application of the method.
(Author/GC)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

INSTITUTE FOR RESEARCH ON POVERTY

REPRINT SERIES

ED155271

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

STATISTICAL DATA IN SCHOOL COURT CASES: AN INTERCHANGE

Richard A. Berk, Anita A. Summers
and Barbara L. Wolfe



00018000

Reprint 258

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Inst. for
Research on Poverty

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC) AND
USERS OF THE ERIC SYSTEM."

Institute for Research on Poverty University of Wisconsin-Madison

Director, Irwin Garfinkel

Assistant Director, Aage B. Sørensen

The Institute for Research on Poverty was established in 1966, by the Office of Economic Opportunity, as a national university-based center for the study of poverty and policies aimed at its elimination. It was established in 1966 by a grant from the Office of Economic Opportunity to the University of Wisconsin. Since 1974 its primary sponsor and major funding source has been the Department of Health, Education, and Welfare, with which it maintains close contact.

The multidisciplinary research staff at the Institute includes those who hold regular teaching appointments at the University of Wisconsin and divide their time between teaching and research, as well as fulltime investigators appointed on a limited-term basis. The senior staff is assisted by junior staff at the graduate level.

The institute offers researchers wide opportunity for interchange of ideas, and provides maximum freedom and facilitating service for poverty-related basic research as well as the study of policy effectiveness.

Research Committee

Peter K. Eisinger (Political Science)
Ross A. Evans (Psychology)
Irwin Garfinkel (Social Work)
Joel F. Handler (Law)
Robert H. Haveman (Economics)
Michael Olneck (Ed. Policy Studies)
Eugene Smolensky (Economics)
Aage B. Sørensen (Sociology)
Karl E. Taeuber (Sociology)
Burton A. Weisbrod (Economics)

MAY 24 1978

STATISTICAL DATA IN SCHOOL COURT CASES: AN INTERCHANGE

Richard A. Berk, Anita A. Summers
and Barbara L. Wolfe

Reprinted by permission from *Journal of Human Resources*
(Summer 1976) © 1976 by the Regents of the University
of Wisconsin System and from *Journal of Human Resources*
(Summer 1977) © 1977 by the Regents of the University
of Wisconsin System.

INTRADISTRICT DISTRIBUTION OF SCHOOL INPUTS TO THE DISADVANTAGED: EVIDENCE FOR THE COURTS

ANITA A. SUMMERS
BARBARA L. WOLFE

ABSTRACT

This paper details a method for those interested in investigating the equity aspects of intradistrict distributions of educational inputs. It should be viewed as a procedure for post-*Hobson v. Hansen* lawyers and for other constituents that steps away from some of the speculative quality of the evidence that has been, and still is, being used. Using detailed individual school data in a systematic way allows the investigator to delineate which unequally distributed inputs positively contribute to stated equity objectives, which inputs are outside the school administration's control, and what the overall equity picture is, starting from the components. The method is applied to data for the Philadelphia school district.

Much as been written on the limitations of the statistical analysis of the now famous *Hobson v. Hansen* cases¹ in Washington, D.C.—cases that involved allegations of racial and income intradistrict discrimination in the allocation of school resources. In this article we detail a method for investigating the equity aspects of intradistrict distributions of educational inputs. This should also be viewed as a blueprint for post-*Hobson v. Hansen* lawyers that steps away from some of the speculative quality of the evidence that has been used. And, it might be viewed as contributing to an agenda for altering some of the institutional arrangements affecting the distribution of educational inputs.

The authors are, respectively, Senior Economist, Federal Reserve Bank of Philadelphia; and Consulting Economist, Federal Reserve Bank of Philadelphia, and Assistant Professor of Economics, Bryn Mawr College.

- * The views expressed do not necessarily reflect those of the Federal Reserve Bank of Philadelphia. [Manuscript received July 1974; accepted October 1975.]
- 1 *Hobson v. Hansen*, 269 F.Supp. 401 (D.D.C. 1967), *affd. sub nom. Smuck v. Hobson*, 408 F.2d 175 (D.C. Cir. 1969), *Hobson v. Hansen*, 327 F.Supp. 844 (D.D.C. 1971). See several discussions of *Hobson v. Hansen* in the Summer 1972 issue of the *Journal of Human Resources*.

I. INPUT DISTRIBUTION AND EXPENDITURE EQUALIZATION

In the second round of *Hobson v. Hansen*, Judge J. Skelly Wright ruled that, for all elementary schools in the Washington, D.C., school district, "per pupil expenditures for all teachers' salaries and benefits . . . in any single elementary school . . . shall not deviate by more than 5 percent from the mean per pupil expenditure for all teachers' salaries and benefits at all elementary schools . . ."² It is not surprising, however, that when plaintiffs use aggregate and few measurements (per pupil expenditures, for example), defendants and the courts will have little difficulty in establishing that "legitimate" differences among schools exist—that is, that differences arising from reasons other than racial or income discrimination will necessarily result in differences in an aggregate measure. Criticisms have been mainly leveled at the outdated income data used to characterize the schools and the nondiscriminatory reasons why salary expenditures might vary by more than 5 percent. It seems to us that more imperative than dissecting the issue of the relationship between size of school and expenditures per pupil³ (a major explanation given for nondiscriminatory variation in expenditures) is increasing the number of inputs that can be examined in a systematic way in an urban school district.

In a recent on-going case involving intradistrict equalization, *Brown v. Chicago Board of Education*,⁴ expenditures are still being looked at in terms of aggregate expenditures, though a cursory swipe has been made at some detail. Wherever individual school data exist,⁵ much deductive testimony can be readily replaced by something closer to the facts.⁶

2 327 F.Supp. 844, at 863.

3 Both plaintiffs ("A Research Report for Plaintiffs," Appendix C to Plaintiff's Memorandum, Dec. 8, 1970, by Stephan Michelson) and defendants ("Defendants' Reply Memorandum to the Court," Jan. 18, 1971, by Dave M. O'Neill, Burton Gray, and Stanley A. Horowitz) addressed themselves to this question.

4 *Dusharn L. Brown, et al. v. Board of Education of the City of Chicago and James F. Redmond*, in U.S. District Court for the Northern District of Illinois Eastern Division, Civil Action No. 71 C 694, *Brief in Support of Plaintiffs' Motion for Preliminary Injunction*, June 28, 1971.

5 These data frequently are not publicly available.

6 *Brown v. Chicago Board of Education*, *supra* fn. 4. In discussing possible causes of differences in per pupil expenditures among schools, the plaintiffs' brief states: "For example, it is conceded that many schools in rapidly growing, and presumably Black neighborhoods have no library facilities at all, whereas schools in stable, White, middle-class neighborhoods, almost invariably, have such facilities. . . . In such situations, it seems apparent that the expenditures for library books will be much greater in the White, middle-class schools." Appropriate examination of the detailed data would allow a decisive description of the relationship between density of black students and library facilities. We found, in Philadelphia, that the distribution of library resources was unrelated to the distribution of black or low-income pupils, with one exception—more library books per pupil were available in elementary schools with more low-income students.

To illustrate the benefits and efficacy of a systematic and detailed examination of the distribution of school inputs, we have looked at the school district of Philadelphia—that is, we have looked at the relationship between the distribution among Philadelphia's public schools of a large number of inputs (including measurements of the quality of teachers and principals) and the distribution among the schools of the proportion of blacks and the proportion of low-income pupils. In Section II we indicate the limitations in considering equalization of educational opportunity issues from the input distribution aspect alone, no matter how detailed the input categories. In Section III, the data and statistical procedures are described. In Section IV the empirical results for the elementary schools of the Philadelphia school district are laid out and summarized. Section V is a statement of what might be learned from the input distribution procedure described in the article.

II. INPUT DISTRIBUTION AND EQUALITY OF OPPORTUNITY

Even a close look at the distribution of inputs to the disadvantaged cannot resolve the question of whether educational opportunity is equal for all. It is not at all obvious what the relationship is between input distribution and educational growth.

First, in certain obvious instances, we recognize that equal inputs do not produce equal outputs. More resources are required to educate blind children than to educate sighted children. Similarly, equal inputs for those with socioeconomic disadvantages and those without would not represent equal educational opportunity. A close look at relative input availability to the advantaged and the disadvantaged can only tell one of the state of the system; it does not answer the question of whether or not this state coincides with views of what constitutes equity. To answer this question, two further steps are required: (1) a definition by parents, courts, and/or legislators of what constitutes equity, and (2) a knowledge of what is required to bring about this state of equity.

A second issue, therefore, is that, even if agreement is reached on what might measure educational opportunity (a minimal goal such as all students increasing their grade equivalent test scores by an amount equal to the number of grades of schooling, or a more aggressive goal of increasing all students who start below grade level to grade level after a specified number of years of schooling?), we need to know what can bring this about. We need to know what package of school inputs is required for each type of student to equip him or her for educational growth. It is these inputs and only these inputs—if they can be identified—whose distribution should be examined to assess a school district's equalization efforts. The so-called education production functions, of which there are many, address themselves to identifying these relationships. But the absence of replicated results, stemming in part from a failure to set up the

empirical investigations in such a way as to allow this, means that the coefficients are not now ready for policy usage.

A third point further limits the inferences that should be drawn about a school district's culpability from data on the distribution of inputs. Some kinds of inputs, it should be recognized, are not controllable by the central administration. Expenditures on plant maintenance are determined annually, but the age of school buildings is not. Urban school administrations are saddled with variations among city schools in the age of the plant facilities and in vandalism expenditures to obtain an equal distribution of plant quality. This unequal input distribution is, presumably, closer to equity objectives than an equal distribution, as is the case with compensatory funds.

Further, the characteristics of the teachers' labor market in urban school districts—an identical wage structure in all schools and seniority transfer rights—mean that “better” teachers are likely to be found in “better” schools. If this outcome is not consistent with equity objectives, then perhaps unions should be joined with superintendents of schools as defendants, since “better” teachers will be in “better” schools as long as union contracts allow for no wage discrimination among different schools. Suing a school administration alone for the consequences flowing from this—more teacher vacancies, better educated teachers in schools with fewer disadvantaged—seems somewhat narrow in aim. (The Philadelphia school district alone, however, would seem to be responsible for not distributing more federal funds to the junior high schools with high proportions of low-income students than to those with low proportions.)

Since we are without adequate knowledge relating to just which inputs affect output, and without a fully defined concept of equality of educational opportunity, parents, courts, and legislators are continuing to look at school inputs to keep tabs. As acceptors of second-best, they are in good company—gross national product accountants value the services of policemen, ministers, and others at input cost, not output value.

But, as acceptors of second-best, it becomes essential to recognize which unequally distributed inputs contribute to achieving equity objectives as they are currently viewed—both horizontal equity (equal treatment of equals) and vertical equity (more inputs to the disadvantaged perhaps)—which resources are outside of a school district's control, and what the overall picture is from looking at all individual inputs separately. Only then is it possible to get some notion of the consistency of the distribution of inputs with equity objectives, and to put the finger on the relevant sinner.

III. THE MEASUREMENT OF INPUT DISTRIBUTION

Our recommended objective in measuring input distribution is to obtain an extensive amount of detailed data on inputs available by individual schools. We

did this for the Philadelphia school district.

We used a combination of physical and dollar measurements of inputs. "Physical" measurements, in many instances, allow one to look at the stock of equipment available (the average number of pupils per lab, for example) and may, therefore, be closer to a quality measurement. Dollar expenditures tell more about the flow of inputs in a time period and, in some instances, may be the best quality measure available (expenditures on remedial education, for example). In several instances we looked at both types of input measurement (the number of library books per pupil and the expenditures per pupil on library books, for example). In addition, a unique type of data was available to us—measurements of teacher "quality" in each school. Averages of these for all teachers were calculated for each school in the Philadelphia school district: the step number (equivalent to experience) of each teacher, the number of years of education of each teacher, the Gourman⁷ rating of the undergraduate college each teacher attended, and the score of each teacher on the National Teacher's Common Examination.

These data were obtained from detailed school-by-school operating budgets (which were published by the school district of Philadelphia for two years); from reports on monthly attendance, school facilities, enrollment, achievement test results, and field operations; and from personnel files.⁸

Inputs are classified in three ways: (1) *Intended to Be Neutral*. These are inputs that are intended to be distributed in a manner consistent with an avowed policy of horizontal equity—distributed, that is, in a way that is unrelated to the proportion of disadvantaged pupils. The school district does not, for example, intend that the number of library books per pupil or the number of credits beyond the B.A. of teachers be determined by the proportion of black or low-income pupils. (2) *Intended to Be Compensatory*. These are inputs that are intended to be distributed in a manner consistent with a defined concept of vertical equity. They are intended to go to particular groups of students to "compensate" them for their socioeconomic handicaps. Thus, federal funds are primarily intended to go to the poor, and remedial education to those disadvantaged by minimal preschool motivation and education. (3) *The World as It Is*. There are many characteristics of the school environment over which the school

7 The Gourman rating was published in *The Gourman Report* (Phoenix, Ariz.: The Continuing Education Institute, 1967), written by Jack Gourman. It is a rating based on the undergraduate programs of nearly all colleges and universities in the United States, with information drawn from professional societies, commercial publications, foundations, etc., as well as the institutions themselves. The areas rated include (1) individual departments, (2) administration, (3) faculty (including student/staff ratio and research), (4) student services (including financial aid and honor programs), and (5) general areas such as facilities and alumni support. The Gourman rating is a simple average of all of these.

8 Detail of these sources accompanies the definitions of symbols used in the full set of tables, available from the authors.

administration has, in any immediate sense, little control and impact—the age of the school and the number of disruptive incidents are examples.

We consider it worthwhile to distinguish between different levels of schooling, since input impact is regarded as differing by age level. Compensatory funding, for example, is poured most heavily into elementary school students. Data for 185 elementary schools, 34 junior high schools, and 22 senior high schools in Philadelphia are analyzed. Linear regressions are used to relate the distribution of resources to the proportion of black students and to the proportion of low-income students.⁹ The relationship is explored through the use of the regressions defined by equations (1) and (2):

$$(1) \quad Z_i^j = \alpha^j \chi_{1i} + \beta^j + u_i^j$$

$$(2) \quad Z_i^j = \gamma^j \chi_{2i} + \delta^j + v_i^j$$

$$i = 1, \dots, N \quad j = 1, \dots, P$$

where Z_i^j refers to the quantity of the j th resource or condition in the i th school, χ_1 is the proportion of black students in the i th school, and χ_{2i} is the proportion of low-income students in the i th school.

The results are classified as to whether or not the direction of the distribution was compensatory (more to the disadvantaged). This classification allows a ready comparison between intent and result.

A neat statistic for summing up all of the individual findings would simplify the problem of drawing conclusions as to whether "discrimination" exists and equity is being served. But no such overall summary statistic exists, and the conclusions must flow from the statistical significance of separate calculations and a judgment about the weight of evidence.

IV. INPUT DISTRIBUTION: THE FACTS

Interpretation of Tables

The results for elementary schools are laid out in Tables 1 and 2.¹⁰ The code for

9 The proportion of low-income students in school is defined as the proportion of students eligible for the free lunch program, which was the basis for the allocation of most of the federally funded programs. We have, in addition, calculated the low-income regressions using school feeder area income estimates. These estimates were derived by a weighted average of income estimates for the blocks in each feeder area. The block income estimates, in turn, were calculated using a procedure combining block housing data and tract income distribution data. The procedure is fully explained, with an accompanying computer program package, in a manual available from the Federal Reserve Bank of Philadelphia.

10 Similar sets of tables describing junior and senior high school input distribution are available from the authors.

TABLE 1
ELEMENTARY SCHOOL INPUTS AND CONDITIONS IN PHILADELPHIA, 1970-71, IN RELATION TO THE
DISTRIBUTION OF BLACK PUPILS $Z_i^j = \alpha^j X_{ii} + \beta^j + u_i^j$

Variable (1)	$\bar{X}$ (2)	σ (3)	Compensatory ^a (4)	Counter- Compensatory ^a (5)	0 to 100% Blacks ^b (6)	(6) as % of $\bar{X}$ (7)	R^2 (8)
<i>Inputs Intended to Be Neutral</i>							
PUPCOST	\$ 536.24	\$ 93.12	+2.250*		\$ +38.49	7.18	.022
SALARY	\$ 463.71	\$ 78.04	+2.607*		\$ +37.19	8.02	.031
CLERICAL	\$ 42.55	\$ 11.27	+2.104*		\$ +4.37	10.27	.018
NONTEACH	279.24	180.83		+0.757	+25.43	9.11	†
PUPRATIO	27.75	3.97	-4.757*		-3.32	11.96	.105
KINDER	\$ 31.13	\$ 14.02	+2.234*		\$ +29.94	96.18	.022
GRADES 1-3	\$ 370.87	\$ 80.24	+1.877		\$ +22.61	6.10	.014
GRADE 4	\$ 358.05	\$ 95.43	+0.578		\$ +9.86	2.75	†
GRADE 5	\$ 336.27	\$ 105.13		-1.427	\$ -26.65	7.93	†
GRADE 6	\$ 324.28	\$ 115.62	+0.076		\$ +1.70	0.52	†
LIBRARY	\$ 4.74	\$ 5.14		-0.514	\$ -0.49	10.34	†
BOOKS	8.05	2.92	+1.002		+0.54	6.71	†
PLAY	53.63	50.33		-3.240*	-29.53	55.06	.049
ART	\$ 1.87	\$ 3.74		-1.693	\$ -1.18	63.10	.010
MUSIC	\$ 8.86	\$ 6.78		-2.144*	\$ -2.67	30.14	.019
HEALTH	\$ 9.63	\$ 6.31	+0.245		\$ +0.29	3.01	†
PLANT	\$ 56.51	\$ 16.23		-0.474	\$ -1.43	2.53	†
CONDITION	2.98	0.90		-0.230	0.00	0.00	†
CAPACITY	93.75%	32.44%		+2.267*	+22.22%	23.70	.022
% VACANT	2.63%	3.70%		+3.386*	+1.50%	57.03	.070

EXPTEACH	7.36	1.16	+0.652		0.14	1.90	†
EXAMTEACH	607.09	20.40		-4.370*	-15.52	2.56	.089
EDUCTEACH	1.86	0.15		-2.195*	-0.06	3.32	.020
SCHOOLTEA	406.05	24.82		-7.858*	-30.86	7.60	.248
DEGREES, P	1.02	0.24	+0.404		+0.02	1.96	†
EXPER, P	8.52	7.04		-5.347*	-6.40	75.12	.130
CREDITS, P	2.06	1.65	+1.177		+0.35	16.99	†

Inputs Intended to Be Compensatory

REMEDIAL	\$ 16.00	\$ 9.10	+3.475*		\$ +5.90	36.88	.057
COUNSEL	\$ 16.56	\$ 6.42	+0.900		\$ +1.07	6.46	†
IMPROVE	\$3010.75	\$6907.59	+5.019*		\$+6051.69	201.00	.116
FEDFUNDS	\$ 75.62	\$ 262.19	+4.921		\$ +91.39	120.85	.112

The World as It Is

DATE	1933	22.92	+0.180		+0.77	0.04	†
%ATTEND	85.94%	4.35%	+1.345		+1.91%	2.22	†
%HIGH	6.18%	8.08%		-7.308*	-9.68%	156.63	.222
%LOW	47.23%	21.02%		+10.716*	+32.87%	69.60	.382
DISRUPT	3.07	3.00		+5.864*	+3.00	97.72	.154

* Significant at the .05 level.

† $R^2 < .010$.

^a $t = \hat{\alpha} / \hat{\sigma}_{\hat{\alpha}}$.

^b $\alpha \cdot 100$.

TABLE 2
ELEMENTARY SCHOOL INPUTS AND CONDITIONS IN PHILADELPHIA, 1970-71, IN RELATION TO THE
DISTRIBUTION OF LOW-INCOME PUPILS $Z_i^j = \gamma^j x_{si} + \delta^j + v_i^j$

Variable (1)	$\bar{X}$ (2)	σ (3)	Compensatory ^a (4)	Counter- Compensatory ^a (5)	0 to 100% Low-Income ^b (6)	(6) as % of $\bar{X}$ (7)	R^2 (8)
<i>Inputs Intended to Be Neutral</i>							
PUPCOST	\$ 536.24	\$ 93.12	+3.445*		\$ +68.86	12.84	.046
SALARY	\$ 463.71	\$ 78.04	+3.462*		\$ +57.98	12.50	.056
CLERICAL	\$ 42.55	\$ 11.27	+2.924*		\$ +7.15	16.80	.039
NONTEACH	279.24	180.83	-0.564		-22.57	8.08	†
PUPRATIO	27.75	3.97	-5.865*		-4.73	17.05	.154
KINDER	\$ 31.13	\$ 14.02	+2.392*		\$ +38.11	122.42	.025
GRADES 1-3	\$ 370.87	\$ 80.24	+3.021*		\$ +42.71	11.52	.043
GRADE 4	\$ 358.05	\$ 95.43	+1.099		\$ +24.85	6.94	†
GRADE 5	\$ 336.27	\$ 105.13		-0.775	\$ -17.19	5.11	†
GRADE 6	\$ 324.28	\$ 115.62		-1.444	\$ -37.80	11.66	†
LIBRARY	\$ 4.74	\$ 5.14		-0.676	\$ -0.77	16.24	†
BOOKS	8.05	2.92	+3.397*		+2.13	26.46	.054
PLAY	53.63	50.33		-3.692*	-39.72	74.06	.064
ART	\$ 1.87	\$ 3.74		-0.921	\$ -0.77	41.18	†
MUSIC	\$ 8.86	\$ 6.78		-1.611	\$ -2.40	27.09	†
HEALTH	\$ 9.63	\$ 6.31	+0.163		\$ +0.23	2.39	†
PLANT	\$ 56.51	\$ 16.23	+2.520*		\$ +8.91	15.77	.028
CONDITION	2.98	0.90		-0.706	-0.14	4.70	†
CAPACITY	93.75%	32.44%		+4.202*	+47.43%	50.59	.083
% VACANT	2.63%	3.70%		+4.502*	+2.10%	79.84	.095

EXPTEACH	7.36	1.16	-1.479	-0.38	5.16	.006
EXAMTEACH	607.09	20.40	-2.106*	-9.41	1.55	.018
EDUCTEACH	1.86	0.15	-2.449*	-0.08	4.30	.026
SCHOOLTEA	406.05	24.82	-5.028*	-25.92	6.38	.116
DEGREES, P	1.02	0.24	-0.495	-0.02	1.96	†
EXPER, P	8.52	7.04	-4.066*	-6.07	71.24	.078
CREDITS, P	2.06	1.65	-0.094	-0.03	1.46	†

Inputs Intended to Be Compensatory

REMEDIAL	\$ 16.00	\$ 9.10	+2.814*	\$ +5.75	35.94	.036
COUNSFL	\$ 16.56	\$ 6.42	+1.938	\$ +2.72	16.43	.015
IMPROVE	\$3010.75	\$6907.59	+4.728*	\$+6830.53	226.87	.104
FEDFUNDS	\$ 75.62	\$ 262.19	+6.152*	\$ +131.67	174.12	.167

The World as It Is

DATE	1933	22.92	-0.202	-1.03	0.05	†
% ATTEND	85.94%	4.35%	+0.360	+0.61%	0.71	†
% HIGH	6.18%	8.08%	-8.857*	-13.27%	214.72	.296
% LOW	47.23%	21.02%	+11.482*	+40.76%	86.30	.416
DISRUPT	3.07	3.00	+3.206*	+2.08	67.75	.048

* Significant at the .05 level.

† $R^2 < .010$.

^a $t = \hat{\gamma}/\hat{\sigma}_{\hat{\gamma}}$.

^b $d \cdot 100$.

the variables is listed in the Appendix to this article. The *t*-ratios are specified for each of the inputs, and they are placed, in accordance with the sign and the nature of the input, in the compensatory column (4) or the counter-compensatory column (5). Asterisks indicate the results that are significant at the .05 percent level. The $\bar{X}$, σ , and R^2 are given in columns 2, 3, and 8. The coefficients are expressed in units reflecting the difference between being in a school that had 100 percent black enrollment or 100 percent low-income pupils (column 6).

It is quite appropriate to consider this full range in the case of Philadelphia. In 1970-71, 9.2 percent of the elementary schools had a 1 percent or less black pupil ratio, and 27.6 percent had a black pupil ratio of 99 percent or more. Similarly, 11.4 percent of those schools had 99 percent or more of their pupils classified as low-income, and 26.5 percent had 1 percent or less of their pupils in the low-income category. To give some notion of the impact of these differences, the coefficients are expressed as percentages of the mean of the variable (column 7).

Consider in Table 1, for example, the question of the distribution of principals' experience (*EXPER*, *P* in column 1) in relation to the distribution of black students at the elementary school level. In column 2, the mean years of experience is seen to be 8.52 years, and in column 3, the standard deviation around the mean is seen to be 7.04 years. The proportion of variability in the experience of principals "explained" by the variation in the proportion of black students, R^2 , is seen in column 8 to be 13 percent. The negative sign of the coefficient is indicated with the sign associated with the *t*-statistic in column 5—that is, the years of experience of the principal decline significantly (indicated by the asterisk) as the proportion of blacks in elementary schools increases. A pupil in a school with all blacks is likely to face a principal whose experience is 6.4 years less (column 6) than a pupil in a school with no blacks. The negative coefficient associated with a 1 percentage point change in the proportion of blacks, $-.0640$, is multiplied by 100 to obtain this value. How "big" is a 6.4 variation? Some notion of this can be obtained by comparing this range with the $\bar{X}$. Thus, in column 7, the entry of 75.12 percent indicates that the variation in principal experience as one moves from an all black to an all nonblack school is substantial—over 75 percent of the mean number of years of principal experience, 8.52 years.

Much is revealed in this disaggregated analysis that is not visible with aggregate measures. Instructional salary cost per pupil (*SALARY*), for example, is an input—when measured in the aggregate—that shows up as being distributed in a compensatory direction (column 4 in Tables 1 and 2). But, when the individual qualities of teachers are examined, the results are counter-compensatory. The greater per pupil salaries in schools with more disadvantaged picks up having fewer pupils per teacher, not having teachers who are more experienced, are from "better" schools, have higher exam scores, or have more education credits beyond the B.A.

It is clear then that school administrators, unlike statisticians, would sigh with relief at seeing a line of daggers ($R^2 < .010$) in the R^2 column for neutrally intended resources and no asterisks with the t -values!

Description of Results

The statistical results suggest that, in 1970-71, the interschool distribution of inputs in Philadelphia had these general characteristics:

1. *The examination scores, education, and quality of undergraduate education were higher for teachers in elementary schools with fewer blacks and low-income pupils.*

At the elementary school level, teacher quality characteristics, where significant, were inferior (in terms of a priori expectations, though not necessarily in terms of impact on educational achievement) in schools with higher proportions of blacks and low-income pupils (see column 5, Tables 1 and 2). These findings also describe the junior high distribution of teacher inputs. At the senior high level, however, these teacher qualities did not generally vary, in important ways, with the proportions of the disadvantaged.

2. *Students in elementary schools where there were higher proportions of black and low-income pupils had significantly less experienced principals* (see column 5, Tables 1 and 2). This was true also for junior high schools.

3. *The proportion of teacher vacancies was higher where there were more disadvantaged students* (see column 5, Tables 1 and 2). The increase in elementary school teacher vacancies between a 100 percent black school and a 100 percent nonblack school was 57 percent of the average number of vacancies—a significant amount. This finding was repeated at both the junior and senior high levels.

4. *Students in elementary schools with higher proportions of black and low-income pupils had a significantly lower ratio of pupils to teacher, very likely due to the greater amount of compensatory education* (see column 4, Tables 1 and 2). Analysis of the junior high level showed the same pattern of distribution.

5. *State and local compensatory funds were directed to the black and low-income student population at the elementary school level* (see column 4, Tables 1 and 2). The finding was similar at the junior high school level, but strongest at the senior high level.

6. *Federal funds were directed to the low-income concentration at the elementary school level.*

Federal funds, though they were clearly distributed in the direction of elementary schools with more blacks, were distributed with great variability (column 8, Table 1). Though something closer to 100 percent of the proportion of variation in federal funds distribution should have been attributable to variation in the density of low-income pupils, this was not the case. Less than 17

percent were (column 8, Table 2, *FEDFUNDS*). Students in schools with 100 percent low-income population compared with students in schools with no low-income pupils did receive 75 percent more than the federal funds average (column 7, Table 2, 174.12 percent)—but the variation was great.

Different findings emerged for the higher levels of schooling. The most disturbing one, perhaps, is that, at the junior high school level, none of the variations from school to school in federal funds per pupil is related to the variation from school to school in the proportion of low-income pupils ($R^2 < .010$). The relevant coefficient was, in fact, negative! At the senior high level, federal funds did go clearly to the schools with higher proportions of low-income students: 45.7 percent of the variation was "explained."

7. *Elementary schools with high proportions of low-income and black pupils tended to be characterized by student bodies with fewer high achievers, more low achievers, and more disruption.*

In the case of elementary school students, concentrations of black and low-income pupils were clearly associated with certain peer group characteristics: the proportion of high-scoring students was very significantly higher in schools with fewer disadvantaged, the proportion of low-scoring students was significantly lower, and the number of disruptive incidents was higher (column 5, Tables 1 and 2).¹¹ Contrary to general belief, the concentrations were not associated, in a statistically significant way, with poorer attendance or older schools (*DATE*) (column 4, Tables 1 and 2). These findings varied somewhat for the higher levels.

Overall, then, neutrally intended resources were distributed in a neutral way in relation to the distribution of elementary school black pupils, except for teacher and principal quality characteristics which were distributed in a counter-compensatory way. They were distributed with a somewhat compensatory bias to elementary school low-income students, except offset again by the distinctly counter-compensatory distribution of the labor input qualities. At the junior high school level, neutrally intended resources were distributed in a somewhat counter-compensatory direction, with teacher and principal qualities again being clearly counter-compensatory in allocation. At the senior high level, the most neutral distribution of neutrally intended resources appears to have occurred. The fact that the high schools draw from much larger geographical areas and, in several cases, from all over the city—therefore, resulting in more heterogeneous student populations—probably accounts for the more neutral distribution.

Finally, it should be noted that if there were a correlation between school size and racial composition or between school size and income composition, and if there were economies of scale, then these results might be misleading. If, for

11 It is not clear whether or not the total number of disruptive incidents in a school or the number of disruptive incidents per pupil is the relevant variable. The results in the tables refer to the former, but calculations on the per pupil basis produce the same conclusions.

example, it were true that schools that were heavily black had relatively large enrollments, and if there were economies associated with size, then spending fewer dollars per pupil on music in such schools might not mean less music education at all. In fact, in the Philadelphia school district in 1970-71, there was virtually no relationship between the proportion of blacks in elementary schools and the size of the school ($r = .133$), nor between the proportion of low-income students and size ($r = -.09$).

V. CONCLUSIONS

What we have tried to do is to recommend and detail a procedure for investigating intradistrict distributions of educational inputs using Philadelphia school district data.

It seems unlikely that the specific results themselves warrant generalization to other areas, even to other large cities. The socioeconomic conditions, from which many pressures that affect educational resource allocation stem, differ. And teacher union strength differs. In comparison with New York, for example, Philadelphia had about the same median family income and about the same proportion of the population below the poverty line, but it had a much higher proportion of black students and a lower median level for education for those aged 25 and over. Union pressures in Philadelphia have not resulted in salaries quite as high as those in New York, but the Philadelphia contract—unlike most others—mandates specific class size. Hopefully, input distribution also depends upon something other than the “squeaky wheel.” Hopefully, it depends upon some more removed judgment about what distribution would be optimal (in terms of achievement growth, for example) for the profile of students in a school district. But there is little evidence to support the proposition that such “rational” distribution occurs.

The specific distributions of inputs should be looked at, however, more fully than they have been. Different conclusions flow from the detail than from the aggregates. A less aggregative analysis would enable courts and constituents to react with fuller knowledge. As it stands now, both react to overall measurements that do not alert interested parties to where the real problems are.

APPENDIX: SYMBOLS USED IN TABLES

<i>PUPCOST</i>	Total cost per pupil (dollars).
<i>SALARY</i>	Instructional salary cost per pupil (dollars).
<i>KINDER</i>	Expenditure per pupil on kindergarten (dollars).
<i>GRADE 6</i>	Expenditure per pupil on basic skills, Grade 6 (dollars).
<i>GRADE 5</i>	Expenditure per pupil on basic skills, Grade 5 (dollars).

<i>GRADE 4</i>	Expenditure per pupil on basic skills, Grade 4 (dollars).
<i>GRADES 1-3</i>	Expenditure per pupil on basic skills, Grades 1-3 (dollars).
<i>LIBRARY</i>	Expenditure per pupil on libraries (dollars).
<i>ART</i>	Expenditure per pupil on art (dollars).
<i>MUSIC</i>	Expenditure per pupil on music (dollars).
<i>HEALTH</i>	Expenditure per pupil on health and physical education (dollars).
<i>REMEDIAL</i>	Expenditure per pupil on remedial education (dollars).
<i>PLANT</i>	Expenditure per pupil on plant operation and maintenance (dollars).
<i>COUNSEL</i>	Expenditure per pupil on counseling (dollars).
<i>CLERICAL</i>	Expenditure per pupil on supervision and clerical (dollars).
<i>IMPROVE</i>	Total expenditure on educational improvement program (dollars).
<i>CONDITION</i>	Condition of main building (scale from 1, worst, to 4, excellent).
<i>DATE</i>	Year in which school was built.
<i>PUPRATIO</i>	Number of pupils per teacher.
<i>NONTEACH</i>	Number of pupils per nonteacher professional.
<i>% VACANT</i>	Percentage of teacher vacancies.
<i>% ATTEND</i>	Average daily attendance (percentage).
<i>% HIGH</i>	Proportion of pupils above 85th percentile in Iowa.
<i>% LOW</i>	Percentage of pupils below 16th percentile in Iowa.
<i>DISRUPT</i>	Number of disruptive incidents per year.
<i>CAPACITY</i>	Number of students attending ÷ pupil capacity (percentage).
<i>LABS, EN</i>	Number of enrolled students per lab.
<i>LABS, AT</i>	Number of attending students per lab.
<i>FEDFUNDS</i>	Total expenditures per pupil from federal funds (dollars).
<i>EXPTEACH</i>	Average step number (related to years of experience) of teachers.
<i>EXAMTEACH</i>	Average score of teachers in National Teacher's Examination, Common.
<i>EDUCTEACH</i>	Average education of teachers (1 for B.A., 2 for M.A., 3 for M.A. + 30, 4 for Ph.D.).
<i>SCHOOLTEA</i>	Average Gourman rating of undergraduate school teachers attended (on a scale from 267 to 770 for Harvard).
<i>DEGREES, P</i>	Number of degrees above B.A. of principal.
<i>EXPER, P</i>	Number of years of experience of principal.
<i>CREDITS, P</i>	Number of additional credits above M.A. + 15 of principal (1 for 15-29, 2 for 30-44, etc.).
<i>BOOKS</i>	Number of library books per pupil.
<i>PLAY</i>	Playground square footage per pupil.

A COMMENT ON SUMMERS AND WOLFE*

As a policy-oriented social scientist, I would be among the first to applaud any effort to improve the quality of data introduced into advocacy proceedings. As the individual primarily responsible for the statistical analysis used in *Brown v. Chicago Board of Education*, I would also support in principle any attempt to assess more accurately the discriminatory allocation of educational resources. However, having participated in numerous legal battles in a variety of roles, I am uneasy with methodological advice implying that statistical procedures can be simply lifted intact from social science applications and placed unchanged into legal settings. In this instance, Summers and Wolfe (in "Intradistrict Distribution of School Inputs to the Disadvantaged: Evidence for the Courts" [5]) neglect the painful reality that almost any set of findings must be subject to vigorous challenge, and subject to challenge under courtroom fact-finding conventions which routinely undermine many popular statistical techniques.

Perhaps the best place to begin is with the question of statistical inference and the ambiguous role that probability plays in legal proceedings. If one employs a frequency conception of probability, inferences must necessarily be placed in the context of a very large number of replications and the proportion of times in the limit that certain phenomena occur. Unfortunately, even when this sampling distribution model is formally appropriate, many practical difficulties remain. For example, while one might wish to determine the "likelihood" of a particular population parameter value from a given set of observations, it can be very misleading to apply a frequency interpretation of probability: either one has correctly estimated a population parameter or one has not. As numerous statisticians have argued, assurances in the aggregate long run have no short-run relevance. Barnett, for instance, observes:

For the present example, what does the confidence interval tell us about θ_0 ? Consider the (precise) 95 percent confidence interval $.241 < \theta_0 < .367$. We must be aware of an appealing misconception! It is *not* true that $.241 < \theta_0 < .367$ holds with the probability .95. Either θ_0 , a determined quantity not a random variable, is in this interval or it is not. All we know

* Whatever merit this comment has must in part be attributed to numerous provocative discussions with Heather Keppler, Robert LeBailly, and Robert Lin during a seminar on uses of inferential statistics. In addition, helpful suggestions were provided by Sarah F. Berk. Finally, much of my recent work on evaluation research has been supported by the Russell Sage Foundation, and this comment is one of the products of that association. [Manuscript received September 1976; accepted October 1976.]

is that in the long run 95 percent of such intervals obtained in similar circumstances to the present one contain θ_0 , 5 percent will not. We have no way of knowing in which category the present interval falls! (Italics are the author's.) [2, p. 35].

Were frequency approaches to statistical inference not sufficiently wanting for many uses within the scientific community, courtroom applications further complicate matters. For example:

Consider first the case in which the existence of a legally significant occurrence or act is itself in question. A barrel falls from the defendant's window onto the plaintiff's head. The question is whether some negligent act or omission by the defendant causes the fall. Proof is available to support a finding that, on over sixty percent of all such barrel-falling incidents, a negligent act or omission was the cause. Should such proof be allowed and, if so, to what effect? [6, p. 1339].

A moment's thought will quickly reveal numerous difficulties with such data in a courtroom setting. For instance, were one to affix a probability interpretation, it is unclear what long-run aggregate occurrences formally communicate about this particular case. Either the defendant was negligent or the defendant was not negligent. Moreover, to what degree can aggregate data from other situations be properly used to make inferences about the case at hand? It would certainly seem unfair to judge the defendant's liability from the actions of other parties or even the defendant's actions in previous situations. In short, besides fundamental limitations with statistical inference in a single instance, legal precedent has yet to resolve the probative weight of probabilistic factual statements.

Given such difficulties with the use of probability in courtroom settings, one can easily cite several serious flaws in the inferential procedures emphasized by Summers and Wolfe. To begin, while it *may* be helpful to establish that the values of certain estimates are unlikely in the long run were their population parameters zero (i.e., statistical significance), interval estimates would seem far more informative. For the case to be won, discriminatory effects must be judged nontrivial,¹ and this requires that the *substantive* impact of legally relevant variables be established along with the role of sampling error. In other words, a binary assessment of statistical significance at some critical level fails to reach the heart of the matter. For probative purposes, one needs to communicate how large a particular set of effects are along with the associated indeterminacy resulting from sampling error.

But more important, the meaning of statistical inference for the reported data is at best ambiguous. On one hand, the authors state their reluctance to

1 De minimis precedents must be formally addressed.

generalize their findings to "even larger cities." Yet, their use of significance tests would seem to imply that some sort of inference to an existing population is being attempted. What is this population? In what sense have we an appropriate (probability) sample? How meaningful is the required sampling distribution? In other words, one would seem hard put to express verbally the evidentiary information being communicated with a claim of statistical significance at the .05 level.²

Still more fundamentally, even if a frequency interpretation could be meaningfully applied to the paper's significance tests, the probative impact would seem problematic. To say that a particular estimate could occur less than 5 percent of the time (i.e., in numerous replications) if the true value were zero, does not formally speak to whether this is one of those instances! Either the given estimate reflects Type I error or it does not.

In short, before social scientists introduce statistical inference into courtroom proceedings, a lot more thought must be exercised. Otherwise, even if we are not misleading ourselves, the impact of our data may be seriously diluted. Our findings will be misunderstood (e.g., statistical significance may be confused with substantive significance) or easily subject to challenge and rendered ineffective.³ Thus, in *Brown v. Chicago Board of Education*, a conscious decision was made not to sample and, hence, to avoid any formal probability statements involving external validity. This may well reflect prudent practice until the courts resolve how statistical inference will be interpreted. In addition, any such rendering of acceptable evidentiary practice will have to resonate with acceptable statistical practice. The latter also has yet to be resolved.

A second group of problems with the suggestions provided by Summers and Wolfe involve the credibility of the substantive models posed and the estimation procedure used. One only has to open any elementary textbook in econometrics to discover easily a number of potential difficulties that would perhaps discredit the paper's conclusions and undermine their intended legal impact. For example, given measurement error in the exogenous variable (even if random), the regression estimates are necessarily biased and inconsistent. Or, one might argue that the residuals for schools in the same (or contiguous) neighborhoods are correlated, and thus significance tests are biased and regression estimates inefficient. Or, one might claim that many of the endogenous variables are

2 Perhaps the authors are not formally addressing external validity, but internal validity. While this is common in econometric models, one still needs to articulate in grounded terms why one should treat the data as a particular realization of some underlying stochastic process and consider the plausibility of thinking in terms of repeated realizations. Similar inferential models borrowed from experimental traditions have analogous ambiguities.

3 The alternative recourse to a Bayesian interpretation of probability eliminates some inferential problems but introduces a host of other difficulties. For example, as a measure of psychological certainty, Bayesian inference is easily subject to challenge as subjective and therefore a less compelling rendering of the "objective" facts of the case.

reciprocally related, and consequently the use of ordinary least squares is inappropriate. I am not arguing that any of these problems are *necessarily* serious; I am only indicating the kinds of challenges that could easily be mounted. The point is that when social scientists attempt to introduce their findings into advocacy settings, they must anticipate a wide variety of potential criticisms even if they are unjustified by social science standards. One may rest assured that should one's data be of real import, every conceivable nit will be picked in an attempt to discredit both the data and the researcher.

While the statistical implications of courtroom tactics have yet to be well articulated for social science methodology,⁴ several preliminary recommendations can be made. First, whenever possible, statistical techniques should be kept simple, at least in actual testimony. It is easy to forget that one's analysis must be fully understood by lay audiences and that statistical elegance per se garners few points. In the *Brown* case, we relied on simple contingency tables precisely because we felt that more complicated statistics would not have been well understood and thus more subject to obfuscation by the other side.

Second, while the analysis presented in court should be kept simple, researchers should attempt to anticipate all possible challenges and run the data necessary to refute potential attacks. Again, drawing from the *Brown* case, we aggregated the school data in several different ways (e.g., weighted and unweighted by the number of students in each school) to anticipate a variety of criticisms. We based our case on the simpler unweighted data, but we had the more complicated analysis ready should it be needed.

Third, one should try to use statistical techniques whose underlying assumptions are as simple and plausible as possible. Thus, for example, even though nonparametric inference is typically less powerful, prudence may argue for its use. Defending assumptions about underlying distributions is a tricky business at best.

Fourth, one must carefully consider the fit between the descriptive properties of one's statistical techniques and the language of legal proof. Standardized regression coefficients, for instance, are difficult to attach to any verbal description of cause and effect, while unstandardized coefficients may easily lead to convenient interpretations. Similarly, the notion of "explained variance" is not readily translated into legal jargon and is perhaps best avoided as a critical element in one's arguments.⁵

Finally, when it is unclear which statistical procedures are formally most appropriate, it is essential to run the data using all plausible options. For example, one might employ ordinary least squares along with logit and probit formulations should the critical dependent variable be dichotomous. If each technique produces comparable substantive results, one's vulnerability to

4 For some interesting exceptions, see [1, 3, 7].

5 For other descriptive problems with variance explained measures, see [4].

challenge is markedly reduced. If different modes of analysis produce different findings, one is at least alerted to the need to justify a given approach over others or, in the absence of such arguments, the need to testify in a far more circumspect manner.

In summary, to the degree that social scientists seriously intend their data to impact advocacy proceedings, they must anticipate courtroom conventions for the determination of fact and the kind of challenges that will likely be mounted. Having seen much respectable work by social science standards fully discredited during cross-examination, I am uneasy with methodological recommendations that ignore the almost inevitable tests to which any important analysis will be put. Unless adjustments for such contingencies are made, the otherwise sensible suggestions from authors such as Summers and Wolfe will be thoroughly vitiated in application.

RICHARD A. BERK
University of California, Santa Barbara

REFERENCES

1. D. C. Baldus and W. L. Cole, "Quantitative Proof of Intentional Discrimination in Judicial and Administrative Proceedings," Midwest Political Science Meetings, 1975.
2. Vic Barnett, *Comparative Statistical Inference*, New York: John Wiley & Sons, Inc., 1973.
3. Richard A. Berk, "Discretionary Methodological Decisions in Applied Research," *Sociological Methods and Research*, forthcoming.
4. Richard A. Berk and others, "Descriptive Distortions in Covariance Based Statistics," *Social Science Research* 5 (1976).
5. Anita A. Summers and Barbara L. Wolfe, "Intradistrict Distribution of School Inputs to the Disadvantaged: Evidence for the Courts," *Journal of Human Resources* 11 (Summer 1976): 328-42.
6. Laurence H. Tribe, "Trial by Mathematics: Precision and Ritual in the Legal Process," *Harvard Law Review* 84: 6 (1971).
7. Marvin E. Wolfgang, "The Social Scientist in Court," *Journal of Criminal Law and Criminology* 65: 2 (1974).

Berk's comment takes up two major issues: (1) the use of classical statistical techniques in the legal context, and (2) the presentation of other-than-simple evidence in the legal context. One general remark might be helpful to the dialogue: Our article attempted to indicate the sort of statistical evidence we regarded as relevant for assessing charges of discrimination in the allocation of school resources. It was not a brief; it was written in a form suitable for the *Journal of Human Resources* and its professional audience. So we would most certainly agree with the commentator that a substantially different format and style would be appropriate in the legal context: confidence intervals would be a preferable way to present the estimates, for example.

(1) Berk strongly attacks the applicability of classical statistical techniques to the demands of legal evidence. Certainly there are problems. He correctly stresses that discrimination, viewed classically, is not stochastic. It either did or did not occur, rendering classical probability statements meaningless. Two other problems associated with the blind use of the 95 percent test of statistical significance could be added.

First, the 95 percent estimate refers to the probability of a Type I error, rejecting the null hypothesis when it is correct. In the context of our article, this means finding discrimination when none is present. Type II errors, concluding no discrimination when, in fact, discrimination exists, is of at least equal concern. Normal statistical procedures do not yield estimates of the probability of Type II errors; indeed, no explicit estimate is possible, since the alternative hypothesis is composite. It is crucial that users of statistical evidence realize that failure to reject the null hypothesis is not equivalent to accepting it. The explicit presentation of interval estimates serves to highlight the power of the significance test, permitting the user to get a feel for Type II, as well as Type I, errors. We would certainly agree that this is essential in the legal context; the professional reader could make these calculations readily, of course, from the data in our Tables 1 and 2.

Second, the classical interpretation of individual *t*-statistics becomes more complex when many equations are estimated. If, for example, 20 independent equations, exploring the discriminatory allocation of educational inputs, were estimated, one would expect to find one statistically significant coefficient (indicating discrimination) at the 95 percent level—even if there was no true discrimination. Individual *t*-statistics cannot be utilized in the normal fashion when they are selected from among many regressions. The normal strictures regarding data-mining are even more pertinent in the legal context—which is

why, in our article, conclusions are drawn from the sweep of results, not from individual findings.

However, one should not get the impression that the techniques of classical statistics are valueless in legal proceedings. The credibility of all evidence is evaluated on the basis of the past experience of a large number of trials. The credence attached to eyewitness testimony is conditioned by knowledge of the proportion of times it is known to have been misleading. In the same way, knowledge that a certain statistical procedure yields interval estimates that contain the true value 95 percent of the time permits reasonable judgment about how much weight to give the quantitative results.

Further, confidence interval estimates do permit some separate judgments of the substantive and statistical significance of the relationships—as well as highlighting the level of power of the tests. A statistically significant parameter with a very small point value is not likely to arouse much concern. But a point value that suggests substantial discrimination may be compelling, even if it is so inaccurately estimated as to not be statistically significant.

Though, on several bases, then, a defense can be put forward for the classical approach in the legal context, it is probably not the best way to use statistics in a legal setting; the subjective concept of probability seems more natural. Courts must make decisions on the basis of limited information; the best judgment must be made on the basis of the judge's beliefs. The judicial setting is one of relatively few where an actual statistical decision problem is confronted; the use of Bayesian techniques might well be appropriate. In any event, the Bayesian view of confidence intervals seems clearly appropriate. As Berk indicates, in footnote 3, the Bayesian method avoids many of the logical difficulties he raises with the classical approach. From a Bayesian perspective, it is entirely legitimate to speak of a 95 percent probability that a true parameter lies within an interval. Berk charges, however, that notions of subjective probability are likely to be perceived, somehow, as less legitimate by the courts. This strikes us as unlikely. Notions of subjective probability raise doubts only in the minds of trained scientists. To the layman, statements expressing "80 percent certainty" or "95 percent confident" seem perfectly natural.

(2) Berk takes issue with the presentation of other-than-simple evidence. There are several points that might be raised in a dialogue on this subject.

First, he relates a "number of potential difficulties" that can be discovered by opening "any elementary textbook in econometrics." They are in the textbooks, but it is not at all clear that they are in our study. Measurement error in the exogenous variable is a problem, but we have every reason to believe that the data on class size, teacher experience, etc., are not in error. All the teacher quality variables, for example, are taken from their personnel records. Since salaries are calculated from these, teachers have a strong incentive to have them accurate. Reciprocal relationship with the endogenous variable is potentially a problem in some studies. There is no reason to believe, however, that black

children seek out the schools with more library books or with teachers with higher National Teacher Examination scores. Residuals among neighborhood schools might be correlated; in fact we find this not to be the case.¹ So, on the witness stand, we'd be well prepared to have the nits picked!

Second, Berk raises the point that "even though nonparametric inference is typically less powerful, prudence may argue for its use." But power is minimizing the magnitude of Type II errors—errors of finding no discrimination when there is in fact discrimination. And this is crucial in the legal context. In fact, the use of a weak test will inherently create a systematic bias in favor of one of the positions—since, loosely speaking, the Type II error will be much greater than the Type I error.

Third, an acceptance of simplicity as a key goal of courtroom presentations would lead to incorrect—biased—results. If a court, for example, is interested in looking at the issue of female wage discrimination, it would not be appropriate to look only at the differences in wages between males and females. Other factors, such as differences in training, enter into the explanation of wages—a case of omitted variables. The included and excluded variables will be correlated, in general; the coefficients on the variable representing the differences in wages between the two groups, b , will therefore be a biased estimator of the true coefficients.² Simplicity has a high cost!

Fourth, the major point of our article is to demonstrate that conclusions that flow from aggregated data (simpler to obtain, analyze, and display) can be shown to be incomplete, when the component data are examined. Thus, from the aggregate measure of per pupil expenditures one would conclude that more is spent on instruction in schools with more low-income pupils; but, an examination of the major components (teacher characteristics, many of which are reflected in salaries) indicates that these schools have teachers who have inferior qualifications.

Finally, it should be noted that the position Berk takes in urging simplicity as a (the?) major objective in the courtroom leads many to the conclusion that social science in an adversary setting is not a possibility—that the truth-seeking characteristic of science is incompatible with the adversary legal process. But courts must grapple with issues that only social science research can illuminate. If they do not, it would mean that issues of sex discrimination in wages and

1 This is based upon a comparison of residuals of achievement growth computed for pupils in the Philadelphia elementary schools, 1967/68–1970/71, and aggregated to the school level. See Anita A. Summers and Barbara L. Wolfe, "Do Schools Make a Difference?" *American Economic Review* (forthcoming).

2 Suppose the true model is $Y = XB + \epsilon$, and the X matrix is appropriately divided into two parts, $X = [X_1, X_2]$. Assume we use only X_2 , though $Y = X_1B_1 + X_2B_2 + \epsilon$. The expected value of b_2 is $E(b_2) = b_2 + (X_2'X_2)^{-1}X_2'X_1B_1$. This equals the true value, b_2 , only if X_1 and X_2 are uncorrelated (the efficiency of X_2 would be improved in any case) or if $B_1 = 0$ (which would violate one assumption of the true model).

racial or income discrimination in the distribution of public goods would be resolved in legislatures, not courts. Perhaps other options should be examined: a review board of social scientists appointed by the appropriate professional organization to appraise the evidence presented to the court? The appointment of "social science clerks" (analogous to law clerks) to aid judges in interpreting the evidence? To us, these seem clearly preferable alternatives to biased or less powerful estimates.

Courts have grappled with the complexities of patents (how can a court understand the intricacies of engineering design?), antitrust issues (how can a court sort out the econometric evidence on price collusion?), malpractice suits (how can a court acquire medical knowledge?), and ballistic evidence (how can a court discriminate between conflicting evidence on trajectories?). Each discipline fancies itself the most arcane. But all of them have managed to be viewed by the courts. Why not social science?

ANITA A. SUMMERS

Federal Reserve Bank of Philadelphia

BARBARA L. WOLFE

University of Wisconsin-Madison

Recent Poverty Institute Reprints

- 244. *Categorization, Belief Similarity, and Intergroup Discrimination*, by Vernon L. Allen and David A. Wilder, 1977.
- 245. *New Directions in State Loan Programs for Postsecondary Students*, by W. Lee Hansen and Suzanne C. Feeney, 1977.
- 246. *The Traffic in Legal Services: Lawyer-Seeking Behavior and the Channeling of Clients*, by Jack Ladinsky, 1977.
- 247. *Comments on Taxes, Equity, and Income Distribution*, by Robert J. Lampman, 1977.
- 248. *'Strategic' Wage Goods, Prices, and Inequality*, by Jeffrey G. Williamson, 1977.
- 249. *Two Notes on the Measurement and Trend in Income Inequality*, by Sheldon Danziger and Eugene Smolensky; and by Sheldon Danziger, Robert Haveman, and Eugene Smolensky, 1977.
- 250. *Comparing Utility Functions in Efficiency Terms or, What Kind of Utility Functions Do We Want?* by Burton A. Weisbrod, 1977.
- 251. *Evaluating the Impact of Public Policies on Regional Welfare*, by Robert H. Haveman, 1977.
- 252. *Poverty, Income Distribution, and Social Policy: The Last Decade and the Next*, by Robert H. Haveman, 1977.
- 253. *Public Interest Law Activity and the Litigation of School Finance: An Appraisal of the Equity and Efficiency Consequences*, by Burton A. Weisbrod, 1977.
- 254. *Tax and Welfare Simplification: An Analysis of Distributional and Regional Impacts*, by Sheldon Danziger and Robert Haveman, 1977.
- 255. *The Effect of Public Policies on the Demand for Higher Education*, by John Bishop, 1977.
- 256. *Microeconomic Determinants of Early Retirement: A Cross-Sectional View of White Married Men*, by Joseph F. Quinn, 1977.
- 257. *The Antipoverty Effectiveness of In-Kind Transfers*, by Timothy M. Smeeding, 1977.
- 258. *Statistical Data in School Court Cases: An Interchange*, by Richard A. Berk, Anita A. Summers and Barbara L. Wolfe, 1977.

For a full list of available Institute Reprints and Discussion Papers write:
Publications Department
Institute for Research on Poverty
University of Wisconsin-Madison
Madison, Wisconsin 53706